



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

Video Retrieval using Textual Queries and Image using the content based and Soccer Videos Summarization using Bayesian Network

Nitin P. Ingle, Prof. Sachin Patil

ME Student, Dept. of Computer Engineering, GHRCEM, Savitribai University of Pune, Pune, India

ME Student, Dept. of Computer Engineering, GHRCEM, Savitribai University of Pune, Pune, India

ABSTRACT: The expanded accessibility of broadband associations has as of late prompted an increment in the utilization of Internet webcasting. Record lectures and positioning them on the Web for contact by understudies has turned into a overall pattern at dissimilar colleges. Videos and Texts demonstrated in lecture are nearly identified with substance of the lecture, gives important source for recovering lecture videos and indexing. Text substance may be separated, then analyse and deducted consequently by OCR (Optical Character Recognition) strategies. In this paper, for remedying lapses in the OCR Transcriptions, we investigated two separate systems connected just to unmatched question words. In the first place strategy produces original set of n-gram strings to match the complete OCR Transcriptions. This n-gram incorporate string with an altered separation of 1 character and all conceivable n-gram substrings with no less than 3 characters. Second system for redressing OCR included the word reference of spelling adjustment strategy gave in MS Words. And second module of video summarization i.e. recent years, it has picked up an extensive passion of specialists and in scholastics from dissimilar regions. While distinguishing video occurrence has been the subject of broad study activities as of late, significantly less existing system has considered multi-model data and issues related effectiveness. Start of soccer matches dissimilar uneasy circumstances develop that can't be adequately judged by the ref board of trustees. A framework that checks equitably picture courses of action would forestall not right understandings in view of a few mistakes, or high speed of the occasions. Bayesian networks give a structure for managing this uncertainty using a basic graphical structure also the probability calculus. We propose an efficient structure for analysing and summarization of soccer videos using object based features. The proposed work utilizes the t-cherry junction tree, an exceptionally recent advancement in probabilistic graphical models, to create a compact representation and great approximation intractable model for client's relationships in an interpersonal organization.

KEYWORDS: pervasive computing, lecture videos, automatic text indexing, OCR, lecture video, Summarization, finding, Bayesian network, Web intelligence.

I. INTRODUCTION

In the previous decade, we have seen a drastic increment in the accessibility of on-line academic lecture material. These educational assets can conceivably change the way individuals learn students with any type of disability can upgrade their educational experience, experts can stay aware of late progressions in their field and individuals of all ages can fulfill their hunger for knowledge. In complexity to numerous other informative exercises be that as it may, lecture transforming has as of not long ago delighted in little advantage from the advancement of human language innovation. In spite of the fact that there has been noteworthy exploration controlled to audio indexing and retrieval.

Embedded content in a feature grouping gives important data of foremost essentialness. Messages more often than not show up as logos, subtitles, inscriptions or pennants in the feature grouping. Illustrations of such educational embedded writings can be to a great extent found in the news and other famous TV broadcastings. In spite of the fact that messages give extra data, not every one of them are essential as they may block critical allotments of a video. There are a few contrasts in the middle of news and lecture speech such as an abstract style and an accessible asset. These days, numerous telecast organizations give news cuts a comparing script through online administration. Since we



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

can without much of a stretch form a preparing corpus using this, telecast news retrieval has been a significant concentrate in talked report retrieval territory. In any case, they as of now give watchword hunt administrations using a content search engine focused around the news script.

Dissimilar to news, we can't undoubtedly get a script of lecture speech. In a business education site, an inquiry is performed using a physically fabricated record. Text is a high state having semantic features which has regularly been used for content-based data retrieval. In lecture videos, writings from lecture slides help as a layout for the lecture and are dangerous for knowledge. In this way in the wake of segmenting a video document into a set of key frames, the content finding method will be executed on each one key edge, and the extracted content articles will be further used as a part of content appreciation also slide structure search forms. Mainly, the removed structural metadata can allow more adjustable video browsing and video finding capacities. Speech is a standout amongst the most needed carriers of information in video lectures. Therefore, it is of unique focal point that this information can be required programmed lecture video indexing. Furthermore, the majority of the current lecture speech recognition frameworks in the explored work can't attain to a sufficient recognition result. A lot of literary metadata will be made by using OCR and ASR system, which opens up the content of lecture videos. To empower a sensible access for the client, the delegate keywords are further concentrated from the OCR and ASR results.

For content-based video look, the look lists are made from. Distinctive information assets, including manual annotations, OCR and ASR keywords, universal metadata, etc. nowadays individuals have a tendency to create lecture videos by using multi-scenes position, by which the speaker and his presentation are shown synchronously. This can be accomplished either by showing a video of the speaker and a synchronized slide record, or by applying a condition of the lecture recording framework, for example, tele-Teaching. Anyplace Solution Kit such a framework which conveys two principle parts of the lecture: the principle scene of lecturers which is recorded by using a video cam and the second which catches the desktop of the speaker as machine during the lecture through an edge grabber apparatus. The key advantage of the recent one for a lecturer is the adaptability.

Sports video sharing over dissimilar networks ought to add to quick acceptance and extensive use of multimedia managements around the world, on the grounds that sports video speaks to extensive groups of viewers. Converting of sports video, for example finding of central actions and summary formation makes it possible to convey sports video similarly over networks, for example, the wireless and Internet, since the imperative semantics for the most part include just a tiny percentage of the total content. The approximation of sports video, be that as it may, drops meaningfully after a reasonably short time of time. So, sports video control essentials to be finished automatically, otherwise, its size, in open, or close ongoing, and the behaviour results must be semantically important. As a standout between the most important components for video information organization systems, the primary functionality of event finding is to extract event and adventure their relationships inside general scale groups. The method has a lot of area wise applications counting video remark and checking, video highlight extraction, content summarization, and many more. The large size of informal organizations focuses the need for exactness in quantification socially with clients that can be updated efficiently. So, the joint distribution characterization of all conceivable social relationships in an informal organization is incomprehensibly, and demonstrating an inference would be computationally not feasible. To focus on these difficulties, we propose a structure expanding on a recently created technique in graphical models to focus the joint distribution. The t-cherry tree was proposed. A t-cherry tree is a structure such that has it has guarantees of precision value of approximation, and taking into consideration exact inference as it is constructed. Various graphical models are often used, for example, graphs [3], but these offers no guarantees on their approximation also may not merge when performing inference [4].

We create a system using t-cherry junction tree to evaluate the basic in a soccer game system. In our work, t-cherry junction tree is constructed by parallelizing most of the computation work. Also, we have present and extract a few events of soccer videos with the features of t-cherry tree. The important method for the current event discovery approaches can be by and big parted into two important steps 1) making video content picture, where the video possessions are removed from basic preparation, and 2) procedure of decision making for finding. In the second step, dissimilar sorts of data mining or arithmetic methods can be linked as indicators with preselected training samples.

Anyhow, completing exact and strong finding is very difficult and testing task. Sports videos invite a wide scope of assembly of people and are by and large transmission for extended periods. For most persons, a compact shortened different (highlights) often seems, by all accounts, to be better-looking than the full length video. Sports highpoints can be nonexclusive made out of motivating events that may release the client's attention. Though nonexclusive highlights are necessarily effective for relaxed video scanning, area specific (or ordered) highlights will



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

support more modified applications. In our method we have used reflex event finding and summarization in soccer videos. The proposed method include algorithm for shot boundary finding, feature extraction, shots1 organization and structure of Bayesian network using t-cherry tree. The presentation of organization is considered using soccer events in some videos. We are detecting seven various events like, goal attempt, foul, offside, goal, corner, card, and non-highlights. Our result shows the better-quality result. The outstanding paper is prearranged in the following manner. In section II we conversed about the related work done by the examiners for event finding in much area. In section III we discussed about the implementation details of the proposed system. In this section we discussed about the system overview, algorithms and mathematical model of the proposed system. In section IV discussed about the result and discussion of the proposed system. In section V we discussed the conclusion of the proposed system and finally we discussed the references used for the paper.

II. RELATED WORK

A texture-based organization for noticing letters in images was showed by K. I. Kim, K. Jung, and J. H. Kim. The outline inspects the textural possessions of works in images using a SVM and spots the content areas by in work CAMSHIFT on the texture organization results. The proposed system can inspire fast content finding, notwithstanding the fact that it doesn't accept the sort of media or the shade and textural properties of writings and is reasonably indifferent to image resolve. It moreover functions estimably in focused writings from mind confusing and textured basics and was found to create a greater implementation than some unlike methods. However, the texture classifier did experience issues arranging little content or content with a low [1].

The algorithm first joins the edge and experiential proofs to focus the confident content lines and after that knows these applicants by operating SVM. The algorithm represented in this paper does not utilize shade, although numerous frameworks moreover make use of color data in detecting content in color images. The principle reason is that the begin purpose of our framework, the edge proof, is basically creating from force in layered image. Varying the RGBcolor image to YUV shade space and performance edge finding in U or V image can without much of a stretch discover this. No worldly data is used as a part of our algorithm. Since content may have diverse developments in feature, content identification is typically performed before following the content among the feature frames. The algorithm exhibited here attains to high identification rate and additionally low false caution rates[2,3]. In quick content line extraction express, this algorithm is quicker than (or equal to) other quick content identification systems, despite the fact that the entire identification methodology is more CPU serious than region-based and edge-based routines[4].

The assessment paradigm of the identification result introduced in this paper is on the premise of right benchmark restriction. This foundation is stricter than complete spread paradigm used as a part of past studies. With this basis, we can quantify the identification execution correctly without needing to demonstrate the last character distinguishes result [5]. In this study, additional Farsi/Arabic text finding and localization procedure is proposed. Initially, with the support of edge extraction, fake angles are learned and text dimension approximation is done. Second, by joining discrete cosine transform constants, texture force based picture is made. Later, another Local Binary Pattern (LBP) picture is acquainted with depicts the acquired texture design[6]. The info picture is then divided into instruction squares and a few features are removed from them and continued into Support Vector Machine (SVM) classifier to sort them into text and non-text groups. Trial results exhibition that the proposed half and half approach can be used as a programmed text finding framework, which is great to text dimension, textual style shade and foundation changeability. Creator exhibit a two stage framework for programmed feature text removal to identify and remove installed feature texts what's more fill-in their remaining districts by fitting information. In the feature text finding stage, text areas in each one frame are discovered through an unsupervised clustering performed on the associated parts created by the stroke width transform (SWT). Since SWT needs a precise edge map, we create a novel edge identifier which profits from the geometric peculiarities uncovered by the band let transform[7,8,9].

The efforts of the text objects of each one frame are poor down to confine feature texts. The known feature text areas are deleted, then, the feature is returned by an in painting plan. The future feature in painting method applies spatiotemporal geometric streams detached by band let's to remake the lost information. A 3d volume regular algorithm, which adventures band let bases in using the anisotropic symmetries, is familiar with complete the in painting scheme [10,11]. The plan does not want additional sequences of action to fulfill visual reliability. E. Leeuwis, M. Federico, and M. Cettolo had focused their exertion on dialect demonstrating. A LM baseline was assessed using unlike sorts of information, which were all faulty, however used as a part of such a path, to the point that their makings



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

were underlined and not their lacks. Using the ITC-irst WSJ AM adjusted on 8h of TED making information, it brought about a WER. D. Lee and G. G. Lee had presented a Korean talked report retrieval framework or secondary school math address structures which uses the material data. From ASR yields and address note, they constructed the general transformed list table, the substance table and the matching table. These tables are used for figuring the significance score. In the study [12] by J. Glass, T. J. Hazen, L. Hetherington, and C. Wang, they closed that the specialized dialect of scholastic addresses and absence of in-area talked information for preparing makes address interpretation a critical test that will require new systems for inferring a vocabulary and dialect model. Pong et al. [8] proposed the algorithm of segmentation in their work is focused around the dissimilarial degree of text and foundation areas. Using limits they endeavor to catch the slide move. The studies depicted in Repp et al. [9] are focused around out-of-the-container business discourse dissimilar programming. Concerning such business programming, to accomplish fulfilling results for an extraordinary working space an adaption methodology is frequently needed, yet the custom expansion is once in a while conceivable[13].

The study of sports video data has gotten great consideration in the last few years but the important interests have been focused on automatic highlight finding, since learning the cuts of highlights actually from a lot of video data is an killing also boring task. Many experts have published papers about soccer video examination and detached event finding. They are based on the remark that the most important events are for the most part taken after by reasonable motion replay. Some single patterns of movielike features can be used to locate objects or other important events. Notwithstanding a lot of examination efforts for soccer summarization in broadcast video, ongoing examination of soccer related images for particular event finding has not gotten great attention in works. The high velocity of soccer events forces strict continuing limitations that normally make the summarization algorithm for these applications unsuitable. Furthermore the finding of high velocity events obliges cams with higher casing rates than the broadcast ones. In these cases the making times must be short to take after the continuous events. Recently, the Bayesian network (BN) has been requested semantic check [14]. In Sun et al. utilizes Bayesian Network for event finding in soccer game videos based on using six various low-level feature including face, sound, gate, texture, caption, and text. Shih et al. create the multilevel semantic network (MSN) to focus the highlights in video of baseball. Additional focus finding method activities visual signals estimated from the video stream, the currently confined playfield zone, player's position, and the shades of players' outfits[15].

The low-level features are used for semantic examination to recognize the highlight, i.e., object, shade and texture features are used to focus the highlight. Xu et al. proposed an effective algorithm for soccer game video, which estimate the play-breaks in soccer game by motion and color features. It notices and tracks useful actions, for example, ball ownership in soccer video that is exceptionally associated to the cam's field-view. Guide based video investigation and indexing systems using the mixture of object descriptors and cinematic are proposed. A video classifying of content-based method focused on announced sports videos using cam motion parameters has been formed. A mixture of the dissertation band energy tracking in sound area and the shade strength pattern recognition in video space gives a valued work for event finding in football game video. A plan for knowledge-based semantic infers for recognition of events in game video has been depicted by a three-tier semantic plan[16].

The (DBN) Dynamic Bayesian network is like a BN and their delays; it tries to bind together time-based quantity with ambiguity. DBN is a valued tool for on behalf of complex stochastic methods. Recent growths in reading and adapting in DBN have been connected to many real world applications. In they propose a robust varying media feature extraction of plan, finding of text and recognition method. The system gives reflex indexing of sports videos based on discourse and video examination. They concentrate on the use of DBN and prove how they can be efficiently linked for joining the authorization got from different media information sources. The BN encodes the conditional reliance relations among a set of arbitrary variables as a diagram. A linkage between two nodes denotes a need in relation which is computed by a conditional probability system.

A graph structure encodes the area knowledge, for example, the relationship between the opinion nodes also the hidden levels, while the many parameters of a conditional probability system can be increased from training data. Although recent methodologies have started to introduce a mixture technique for video explanation, the state-of-the-art successes in sports video explanation still experience the ill effects of two important disadvantages: 1) a definitive degree of events finding and explanation (i.e., where to start and complete the extraction) and 2) the absent of a general set of features for noticing dissimilar events and sports. For instance, Ekin et al. have optional that independent finding ought to be checked within the video outlines that are between the universal cam shot that causes the objective and the universal shot that demonstrates the restart of the enjoyment. Nevertheless, this extent of finding template was not

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

reliably used for other events. Similarly, Han et al. used a static temporal segment of 3040's which may cut the semantic stream content. With respect to the absence of general features set, the particular features that best show a highlight event are often selected using area knowledge. For instance, the whistle in soccer is used for detecting foul and offside, while pleasure also objective range are used for recognizing objective attempts. In the previous study, object color and texture features are used to generate highlights and also parsing of TV soccer game programs [17,18]. Use of object motion lines and connections for football are play for organization and also for soccer game finding of event. On the other hand, depend on pre-extracted accurate object lines, which were obtained physically in; subsequently, they are not practical for constant applications.

III. PROPOSED ALGORITHM

1. System Architecture of content based

Following Figure 1 shows proposed system architecture. The detailed description is as follows:

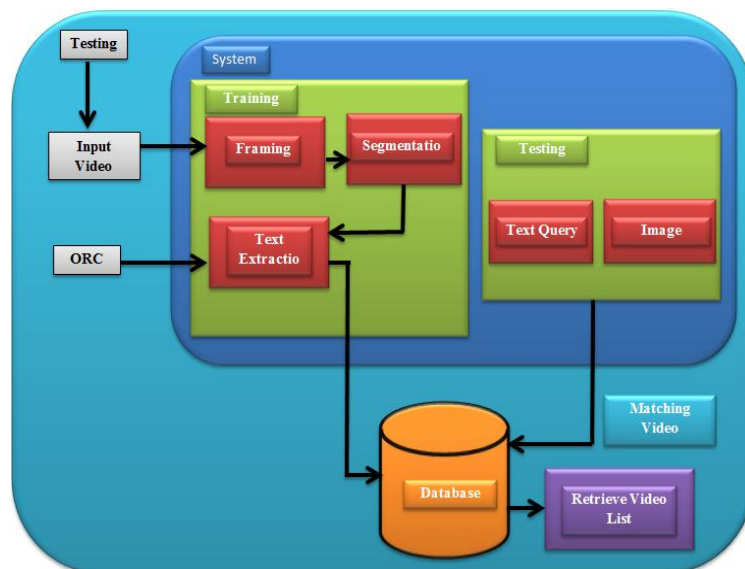


Fig.1. System Architecture of content based Video Retrieval

Working of the system is divided into two parts:

- 1) Training Phase
- 2) Testing Phase

Training Phase: We can take numbers of videos in for training process as input. At training phase initially we train videos one by one, dissimilar frames are extracted from selected video and segmentation of this frames are performed.

OCR: Optical Character Recognition is applied on the extracted frames. i.e. we extract the text from each frame, after extracting text we can check this text for spelling. Correction of unmatched word can be done in this phase. We are also retrieving text from audio by applying Automatic Speech Recognition technique. Indexing is performed on Text extracted from ASR and OCR and this indexing values are further stored in dictionary. It is used to compare it with user query.

Testing Phase: In this phase user enters the input query text ie Text for video retrieval. If input query is incorrect we can correct it by applying n gram technique. Indexing of this text is formed and these indexing values are compared with information stored in directory and at the end related videos are retrieved.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

2. Architecture of video summarization using soccer's and Bayesian network

In the Second architecture of video summarization using soccer's and Bayesian network firstly we give in the fig. 2. Number of videos to the system as inputs. In Boundary and Replay finding features are extracted. From various videos frames are formed. Group is essential to classify frames into boundary and non-boundary finding. Information about the shot viewpoint type can ready to cooperative reminders on semantic content of the video. By and large, four classes of shot viewpoints can be considered in soccer videos: long view, medium viewpoint, close-up viewpoint, and out-of-field viewpoint. Replays are important units in sports video semantic examination. They are used to give more information and details on happened events. Another semantic unit is introduced for sports videos called play-break procedure. Each one play-break procedure consists of a few shots. These semantic units are considered as the smallest (first level of) semantic unit in sports and observation videos. In soccer videos, the change is in a play mode when the ball is in the field and the game is continue; a break mode is the match situated, like, at whatsoever situation the game is stopped due to event of a game (e.g., foul, corner).



Fig.2. System Architecture of Video Summarization

These features are absolutely the events then again concepts of the video. In soccer videos, a few events are considered as high state features. The important feature of a Bayesian network is its ability to capture conditions among extracted features. In fact, extracted features are considered as the unequal variables of the network. There are a few methods for defining the structure. We have used t-cherry for structure estimation of the Bayesian network. In the proposed system, a few events are known using a Bayesian network. It is an acyclic directed graph which represents conditions situated among a set of unsystematic variables. As the Bayesian network is a compelling tool for learning complex examples, a Bayesian network-based system is proposed for finding and summarization in soccer features. The important part of the system is building the Bayesian network, for which the structure is assessed using the t-cherry



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

tree. The Bayesian network comprises of only one variable and a few unsystematic variables which are observed. Network structure is controlled by the t-cherry. We have proposed a novel system for figuring the joint disseminations of continuous unsystematic variables in the Bayesian network, using the Copula hypothesis. The joint distributions of unsystematic variables of the network are displayed by applying the Farlie-Gumbel-Morgenstern group of Copulas. The standard meaning of a copula is a multivariate aggregate distribution capacity characterized on the unit shape [0, 1]n, with consistently disseminated minor distributions. We can utilize Copula for discovering the joint distribution of a few random variables.

3. Technical Specification

A.1. Algorithm/ Technique of Content Based

All non-title text line objects are further classified into three classes: content text, key-point and foot line. The organization is based on the height and the average stroke width of the text line object, which is described as follows:

key-point if $S_t > S_h$ mean $t > h_{mean}$

footline if $S_t < S_h$ mean $t < H_y$ mean = Y_{max}

content text otherwise; where S_{mean} and H_{mean} denote the average stroke width and the average text line height of a slide frame, and Y_{max} denotes the maximum vertical position of a text line object.

Keyword extraction and Video search

Formula for calculating TFIDF score

$$tf\ id_{seg} - internal(k_w) = 1/N(tf\ id_{ocr} \ 1/n_{type} \sum n_{type} \ i \ 1wi + tf\ id_{asr} \cdot w_{asr}) \dots \dots (1)$$

Where k_w is the current keyword, $tfidfor$ and $tfidfar$. denote its TFIDF score calculated from OCR and ASR resource respectively, w is the allowance issue for different properties, n_{type} means the number of various OCR text line types. N is the number of obtainable information resources, in which the current keyword can be found, namely the consistent TFIDF score does not equal 0.

A.2 Algorithm video summarization using soccer's and Bayesian network

Algorithm for t-cherry tree

- It is used for formation of the Bayesian network structure.
- It is generalization of the Chow-liu tree.
- It is K-order tree.

C=set of cluster

S=separator

There are two phases for construction of tree.

1) Construction of table

2) Cluster addition

First, a table T listing all of the k possible cluster-separator pairs is constructed.

C' = cluster-separator pair entry, $C \setminus S'$ is called the dominating vertex of the cluster Calculate weight of entry as,

$$\omega = I(X_{C'}) - I(X_{S'})$$

Tables sorted by heaviest weight. After table construction cluster addition phase takes place. Starting with an empty junction tree, determine whether the heaviest remaining entry can be added to the junction tree. To facilitate checking if the cluster-separator addition is valid, maintain a binary vector of all nodes currently represented in the junction tree. This vector is denoted V, where $V(v) = 1$ represents that variable v is contained in at least one cluster in the junction tree. Next, examine the condition under which a cluster separator pair can be added and maintain a valid t-cherry Junction tree.

B.1. Mathematical Model of content based

Let, the system S is represented as:

$$S = \{T, F, S, T, C, R\} \dots \dots \dots (2)$$

1. Training input video

T is a set of all training input videos given to the system,



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

$T = \{t_1, t_2, t_3, \dots\}$ Where, $t_1, t_2 \dots$ are the number of input videos given.

2. Framing

F is a set of framing input videos which are extracted

$F = \{f_1, f_2, f_3, \dots\}$ Where, $f_1, f_2 \dots$ are the number of dissimilar frames.

3. Segmentation Phase

Let, S is a set of segments

$S = \{s_1, s_2, s_3, \dots\}$ Where, $s_1, s_2, s_3 \dots$ are the number of dissimilar segments

4. Text Extraction

Let, T is a set of extracting text

$T = \{t_1, t_2, t_3, \dots\}$ Where, $t_1, t_2, t_3 \dots$ are the number of extracted texts.

5. Checking for Spelling

Let, C is a set for spelling check

$C = \{c_1, c_2, c_3, \dots\}$ Where, $c_1, c_2 \dots$ are the number of checker for spelling.

6. Retrieved Video Output

Let, R is a set for retrieve videos

$R = \{r_1, r_2, r_3, \dots\}$ Where, $r_1, r_2 \dots$ are the number of various retrieved videos as output.

B.2. Mathematical Model of video summarization using soccer's and Bayesian network

Let, System S is represented as: $S = \{D, F, E, T, N\}$

1. Boundary and Replay finding :

Consider, D is a set of detecting shot boundary and replay

$D = \{d_1, d_2, \dots\}$ Where, d_1, d_2, \dots are number of findings.

2. Feature Extraction:

Let F is a set for feature extraction

$F = \{f_1, f_2, f_3, \dots\}$ where, f_1, f_2, \dots are number of features.

3. Structure Estimation:

Let, T is a set for structure estimation forming a tree,

$T = \{t_1, t_2, t_3, \dots\}$ where, t_1, t_2, \dots are number of structure form.

According to Bayesian rule:

$$P(A|B) = (P(A,B))/P(B)$$

For this Bayesian network is used

$$FXY(x, y) = C(FX(x), FY(y))$$

The joint distribution of x and y is distribution function. It is copula function.

4. Event Finding:

Let, E is a set for event Finding.

$E = \{e_1, e_2, e_3, \dots\}$ where, e_1, e_2 are the number of event used for finding.

5. Video Summarization:

Let, N is a set for summarization for various shots

$$I. N = \{N_1, N_2, \dots\} \text{ WHERE, } N_1, N_2, \dots \text{ ARE NUMBER OF SUMMARY OUTPUT FOR VARIOUS INPUT.}$$

IV. SIMULATION RESULTS

The organization is constructed using Java framework on Windows stand. The Netbeans is used as a development tool. The system doesn't require any specific hardware to run; any standard machine is capable of running the application.

1. Content Based

A. Dataset

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

We will use dissimilar video lectures which are done online in future. Videos lectures can also be downloaded from various standard websites available on web.

B. Results

In following Table I show precision, recall and f1 measure value. Various setups are considered for that. This bar graph shows the accuracy Evaluation of Task1 of the existing system, accuracy is measured by Recall, precision and F1 Measure.

Setup	Recall	Precision	F1 Measure
Key Frame & Video	0.99	1	0.99
Keyword & Video	0.99	1	0.99
All Feature & Video	0.96	0.99	0.97
Outline & Video	0.87	0.95	0.91
Video Only	0.81	0.83	0.82

Table 1. For Existing System

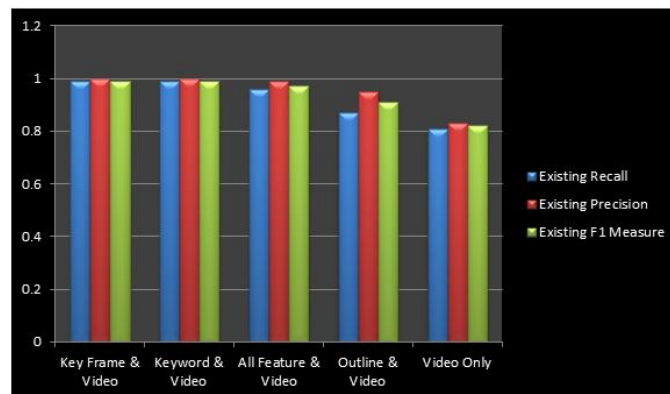


Figure 2. Graph for Existing System

2. Video summarization using soccer's and Bayesian network

A. Dataset

In this work we have used several soccer videos. For random soccer game videos we have taken from any country within no time limit of videos.

B. Results

Following table I shows training time table.

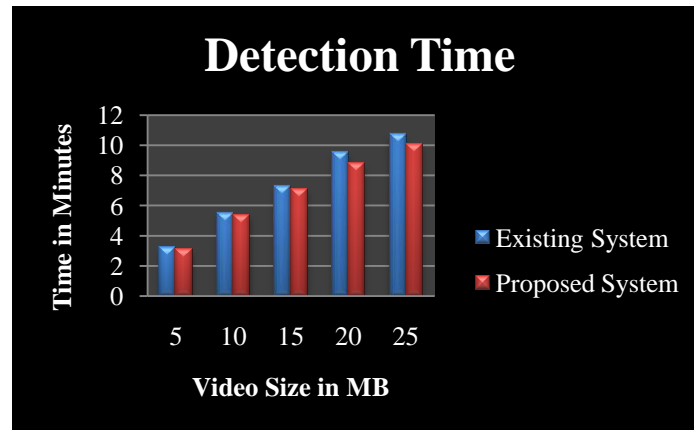
Sr. No.	Video Size in MB	Existing System	Proposed System
1	5	3.2	3.1
2	10	5.5	5.4
3	15	7.3	7.1
4	20	9.2	8.6
5	25	10.7	10.1

Table 2. Finding time in minutes

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015



In the above table II it shows time required for finding of events in both existing system and proposed system. For every dissimilar input the time required in proposed system is always less than existing system. So, our experimental evaluation results shows that proposed system improves accuracy in terms of time utilize to detect an event. Following figure 3 shows graph for finding time according to table II.

V. CONCLUSION AND FUTURE WORK

We displayed a methodology for content based video of lecture indexing and recovery in expansive archives of lecture video. So as to check the examination theory we apply visual and additionally sound asset of address features for focused content based metadata consequently. A few novel indexing features have been created in an expansive video lectures portal by using those metadata and a user study has been led. In our work, we use techniques for correcting errors in OCR transcriptions. This technique first generates n-gram strings for matching OCR unedited transcriptions. Then gram string contains all possibility for substrings having minimum 3 characters.

One of greatest imperative determination for the acceptance of video data is the excellence of its content and its status. The meaning of information detection over huge video groups is important but to be totally studied. The Bayesian network is used for classifier in soccer game for the purpose of event finding. Notwithstanding some earlier methods that are based on framing or shot, our proposed method used the play-break preparation as a unit which extracts additional convincing features from the video also reduces the appreciative making cost. Our principle input was the use of Copula and also t-cherry tree for assessing the joint deliveries in the Bayesian network. Our result shows upgrading than existing system while comparing results. In future one can also focus on issue related to event credit and summarization in football videos.

REFERENCES

- [1] MostafaTavassolipour, MahmoodKarimian, and ShohrehKasaei, "Event Finding and Summarization in Soccer Videos Using Bayesian Network and Copula", *IEEE transactions on circuits and systems for video technology*, Vol.24, No. February 2014.
- [2] E. Kovacs and T. Szantai, "On the approximation of a discrete multivariate probability distribution using the new concept of t-cherry junction tree", vol. 633, pp. 3956, 2010.
- [3] H. Zhuang, J. Tang, W. Tang, T. Lou, A. Chin, and X. Wang, "Actively learning to infer social ties, Data Mining and Knowledge Discovery", vol. 25, no. 2, pp. 270297, 2012.
- [4] KwangKwang In Kim, Keechul Jung, Jin Hyung Kim, "Texture-based approach for text finding in images using support vector machines and continuously adaptive mean shift algorithm", *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (Volume:25, Issue: 12) .
- [5] Datong chen,"Text Identification in complex Background Using SVM", *IEEE*, 2001.
- [6] MohieddinMoradi, SaeedMozaffari, "Hybrid approach for Farsi/Arabic text finding and localization in video frames",*IET Image Process.*, 2013, Vol. 7, Iss. 2, pp. 154–164 doi: 10.1049/iet-ivr.2012.0441.
- [7] Ali Mosleh," Automatic Inpainting Scheme for Video Text Finding and Removal", *IEEE Transaction on mage Processing*, Vol. 22, No. 11, November, 2013.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

- [8] E. Leeuwis, M. Federico, and M. Cettolo, "Language modeling and transcription of the ted corpus lectures," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2003, pp. 232-235.
- [9] D. Lee and G. G. Lee, "Korean spoken document retrieval system for lecture search," in *Proc. ACM Special Interest Group Inf. Retrieval Searching Spontaneous Conversational Speech Workshop*, 2008.
- [10] J. Glass, T. J. Hazen, L. Hetherington, and C. Wang, "Analysis and processing of lecture audio data: Preliminary investigations", in *Proc. HLT-NAACL Workshop Interdisciplinary Approaches Speech Indexing Retrieval*, 2004, pp. 912.
- [11] T.-C. Pong, F. Wang, and C.-W. Ngo, "Structuring low-quality videotaped lectures for cross-reference browsing by video text analysis," *J. Pattern Recog.*, vol. 41, no. 10, pp. 3257-3269, 2008.
- [12] S. Repp, A. Gross, and C. Meinel, "Browsing within lecture videos based on the chain index of speech transcription," *IEEE Trans. Learn. Technol.*, vol. 1, no. 3, pp. 145-156, Jul. 2008.
- [13] Haojin Yang and Christoph Meinel, "Content Based Lecture Video Retrieval Using Speech and Video Text Information", *IEEE Transaction on Learning Technologies*, Vol. 7, No. 2, April-June 2014.
- [14] A. Haubold and J. R. Kender, "Augmented segmentation and visualization for presentation videos," in *Proc. 13th Annu. ACM Int. Conf. Multimedia*, 2005, pp. 51-60.
- [15] D. A. Sadlier, N. O'Connor, S. Marlow, and N. Murphy, "A combined audio-visual contribution to event finding in field sports broadcast video", in *IEEE Int. Symp. Signal Processing and Information Technology, Darmstadt, Germany*, 2003.
- [16] C. Wu, Y.-F. Ma, H.-J. Zhang, and Y.-Z. Zhong, "Events recognition by semantic inference for sports video," in *Proc. IEEE ICME 2002, Lausanne, Switzerland*, 2002, vol. 1, pp. 805-808.
- [17] M. Petkovic, V. Mihajlovic, W. Jonker, and S. Kajan, "Multi-model extraction of highlights from formula 1 programs," in *Proc. IEEE ICME, Lausanne, Switzerland*, 2002.
- [18] D. Yow, B.-L. Yeo, M. Yeung, and B. Liu, "Analysis and presentation of soccer highlights from digital video," in *Proc. Asian Conf. on Comp. Vision (ACCV)*, 1995.
- [19] Y. Gong, L. T. Sin, C. H. Chuan, H.-J. Zhang, and M. Sakauchi, "Automatic parsing of soccer programs," in *Proc. IEEE Int. Conf. Mult. Comput. Syst.*, 1995, pp. 167-174.
- [20] S. Intille and A. Bobick, "Recognizing planned, multi-person action, *Comput. Vis. Image Understand.*", vol. 81, no. 3, pp. 414-445, Mar. 2001.
- [21] V. Tovinkere and R. J. Qian, "Detecting semantic events in soccer games: Toward a complete solution," in *Proc. IEEE Int. Conf. Mult. Expo (ICME)*, Aug. 2001.